

Task 3: Making changes to Studily

Studily is an AI enabled application that helps students who learn from YouTube to not get distracted by watching non-educational video in their recommendation.

How it works:

I used selenium which is a very popular browser automation library in python. Once the application is started, YouTube is opened on a chrome browser and the user is allowed to watch any educational video they wish to see. The program regularly checks the title of the video that the user is watching and uses AI to make a prediction if the title of the video implies Educational or Non-Educational. If it is Non- Educational, the user is redirected to a different URL, thus not allowing them to watch the video.

My task was to make changes to this application:

- a. Add multiple categories, instead of binary, for the models to classify videos under.
- b. Define a minimum confidence threshold for the models to classify videos under - if the target prediction is below the threshold, simply raise an exception “unable to categorize”
- c. List out the probabilities for the top 3/top 5 predicted classes.

To start-off, I needed new data to work with. For this I used my youtube data extraction tool which uses YouTube API to get the titles of videos based on keywords provided.

I gathered around 1500 datapoints for the following classes:

- a. Educational
- b. Music
- c. Sports
- d. Movies
- e. Gaming

```
dataset['Category'].value_counts()
```

```
0    1485
3    1480
4    1479
1    1477
2    1459
```

```
Name: Category, dtype: int64
```

Now, each datapoint had to be cleaned and converted into TF-IDF vectors. But before that stop words like {"the", "I", "a", "to"} etc had to be removed. For this I used the natural language toolkit library, in short NLTK which comes with a list of all the stop words. Looping through each word of a title, the stop words were removed.

Then the data was converted into TF-IDF vectors. This is important because we need to convert the text into some form of meaningful numbers. TF-IDF basically assigns each word some importance value based on how many times a word has appeared and how many times it is there in each document.

Finally, the vectors were fed into a PassiveAggressiveClassifier. This works by changing its weights and biases aggressively when it gets a prediction wrong during training, and remaining passive when it gets the predictions correct.

	precision	recall	f1-score	support
Educational	0.91	0.90	0.91	377
Music	0.97	0.96	0.97	360
Sports	0.90	0.92	0.91	347
Gaming	0.93	0.95	0.94	376
Movies	0.95	0.92	0.93	385
accuracy			0.93	1845
macro avg	0.93	0.93	0.93	1845
weighted avg	0.93	0.93	0.93	1845

Here is the evaluation report

INTEGRATION WITH YOUTUBE AND SELENIUM

Live: IND Vs ENG, ICC World Cup 2023 | Live Match Centre | India Vs England | 1st Innings

Top chat

PROBLEMS OUTPUT TERMINAL PORTS DEBUG CONSOLE


Educational : 15.383166269777565
Music : 10.254423874187935
Sports : 48.00212809146668
Gaming : 11.236098244995969
Movies : 15.124183519571838

Sports 48.00212809146668 % confidence


YouTube

how to play basketball


Sign in




Basic rules of Basketball (Fouls & Violations)
ThingsILearned TV
28K subscribers
Subscribe
27K
Share




How to Shoot a Layup in...
ThingsILearned TV
29K views · 3 years ago
3:58




RULES OF BASKETBALL...
Basketball U
161K views · 2 years ago
19:44



혹시 화났어요?
예편디라
4.8K views · 2 days ago
8:01



How To Play Defense For...
Get Handles...
195K views · 8 months ago
7:58



BASKETBALL RULES EXPLAINED
Rohan Bhaunt

Basic rules of Basketball (Fouls & Violations) - YouTube
['basic rule basketball foul violation']
Educational : 16.47664930558217
Music : 6.573801028850252
Sports : 62.7158990042698
Gaming : 6.206359431497239
Movies : 8.027291229800525

Sports 62.7158990042698 % confidence

Best of Arijit Singhs 2022 | Arijit Singh Hits Songs | Latest Bollywood Songs | Indian songs. - YouTube

youtube.com/watch?v=_z8UyrLIM7k

Chrome is being controlled by automated test software.

Get captions for your audio and video

arijit singh best songs

Sign in

ARIJIT SINGH
— The Best Of —

01. Agar Tum Saath Ho
02. Hawayein
03. Hamari Adhuri Kahani
04. Tera Yaar Hoon Main
05. Ae Dil Hai Mushkil
06. Soch Na Sake
07. Samjhawan
08. Roke Na Ruke Naina
09. Phir Bhi Tumko Chaahunga
10. Channa Mereya
11. Enna Sona
12. Teri Meri Kahaani
13. Raabta
14. Dekha Hazaro Dafa
15. Kabira Encore
16. Phir Kabhi
17. Itni Si Baat Hai
18. Dekh Lena

Best of Arijit Singhs 2022 | Arijit Singh Hits Songs | Latest Bollywood Songs | Indian songs.

MUSIC WORLD
278K subscribers

Subscribe

572K

Share

3X Your Salary
Sponsored · jatan.skill
Sign up

ARIJIT SINGH
The Melodies
Hits of Arijit...
Yuvraj studio (Ufet...
11M views · 1 ye...

ARIJIT SINGH
Best of Arijit
Singhs 2023 ...
izharaar lofi
1.2M views...

NONSTOP LOFI
SLOWED & REVERB
35:06
Rs Lofi lovers
2.7K views · 2 da...
New

Best of Arijit

```
Best of Arijit Singhs 2022 | Arijit Singh Hits Songs | Latest Bollywood Songs | Indian songs. - YouTube
['best arijit singhs arijit singh hit song latest bollywood song indian song']
Educational : 9.478909361751581
Music : 70.90433085994725
Sports : 7.131409176629232
Gaming : 7.4621414665477985
Movies : 5.023209135124135

Music 70.90433085994725 % confidence
```

The program is also capable of making predictions on Titles of different languages such as Hindi, german, French, etc



गणित के 27 सबसे चर्चित टॉपिक !! जहां से सबसे ज्यादा प्रश्न पूछे जाते हैं !! इनको पढ़कर रिजल्ट तय



DREAM COMPE...
761K subscribers

Subscribe

16K



Share



Show chat replay



गणित के 27 सबसे चर्चित टॉपिक !! ज...
DREAM COM...
93K views • 4...



COUNTING
FIGURE II...
DREAM COM...
19K views...



गणित के 27 सबसे चर्चित टॉपिक !! ज...
DREAM COM...
81K views • 4...



LCM तथा HCF का
ONESHOOT...
DREAM COM...
4.2K views...
New

27 most popular topics of mathematics!! From where most questions are asked!! Result decided after reading these - YouTube
['popular topic mathematics question asked result decided reading']
Educational : 43.89238370896999
Music : 15.645053603016734
Sports : 10.46869650587808
Gaming : 13.79243448376414
Movies : 16.20143169837105
Educational 43.89238370896999 % confidence